

AD

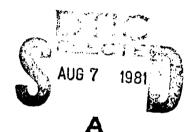
AD A 102540

TECHNICAL REPORT ARBRL-TR-02331

QUANTILES OF MULTI-SAMPLE SMIRNOV TYPE STATISTICS

Mark P. Becker Malcolm S. Taylor

June 1981



C FILE COPY



US ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND BALLISTIC RESEARCH LABORATORY ABERDEEN PROVING GROUND, MARYLAND

Approved for public release; distribution unlimited.

81 8 06 045

Destroy this report when it is no longer needed. Do not return it to the originator.

Secondary distribution of this report by originating or sponsoring activity is prohibited.

Additional copies of this report may be obtained from the National Technical Information Service. U.S. Department of Commerce, Springfield, Virginia 22161.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

The decomplete names or manufacturers' names in this expent to a not conditute indomnement of any commercial product. UNCLASSIFIED SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
TECHNICAL REPORT ARBRIL-TR-02331 D-410)	5 40
4. TITLE (and Subtitie)	5. TYPE OF REPORT & PERIOD COVERED
QUANTILES OF MULTI-SAMPLE SMIRNOV TYPE STATISTICS	5
	6. PERFORMING ORG. REPORT NUMBER
7. Author(a)	B. CONTRACT OR GRANT NUMBER(N)
Mark P. Becker Malcolm S. Taylor	1.11/2
9. PERFORMING ORGANIZATION NAME AND ADDRESS US Army Ballistic Research Laboratory	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
ATTN: DRDAR-BLB	
Aberdeen Proving Ground, MD 21005 11. CONTROLLING OFFICE NAME AND ADDRESS	RDTGE /1L161102AH43
USAARRADCOM Ballistic Research Laboratory	JUNE 1981
ATTN: DRDAR-BL Aberdeen Proving Ground, MD 21005	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(II different from Controlling Office)	15. SECURITY CLASS (of this report)
	UNCLASSIFIED 15m. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)	
Approved for public release; distribution unlimit	od
	eu.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different fr	om Report)
18. SUPPLEMENTARY NOTES	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number Birnbaum-Hall Test Nonnarametric	
Smirnov Type Statistics Nonparametric S Hypothesis Test	statistics ting
Let $X_1^{(1)}, X_2^{(i)}, \dots, X_{n_i}^{(i)}$, $i = 1, 2, \dots, c$, be samp	oles of c independent random
variables $X^{(i)}$ with continuous cdf $F^{(i)}$. To test	the hypothesis H_a : $\Gamma^{(i)}$ =
$F^{(j)}$, i, j = 1,2,, c, Birnbaum and Hall introdu	iced the statistics
$D(n_1, n_2, \ldots, n_c)$ and $D^{T}(n_1, n_2, \ldots, n_c)$ and deter	mined their exact small
sample distributions for $c = 2,3$ and $n_i = n_j$. Thi	s paper considers exact small
DO FORM 1473 FONTION OF 1 NOV 65 IS ORSOLETE	The second second second

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

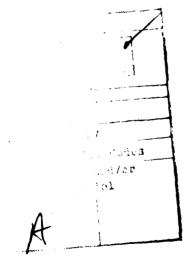
393471

. # Z

UNCLASSIFIED

TABLE OF CONTENTS

		Pa	ge
I.	INTRODUCTION AND SUMMARY		5
II.	A GEOMETRICAL PERSPECTIVE		6
III.	DETERMINATION OF EXACT SMALL SAMPLE DISTRIBUTION		7
IV.	TABLES		7
	DISTRIBUTION LIST		15



I. INTRODUCTION AND SUMMARY

Let $X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)}$, $i = 1, 2, \ldots$, c be samples of c independent random variables $X^{(i)}$ with continuous cumulative distribution functions $F^{(i)}$ and empirical distributions $F^{*(i)}$

$$F^{*(i)}(x) = 0$$
 $x < X_1^{(i)}$
 $= k/n_i$ $X_k^{(i)} \le x < X_{k+1}^{(i)}$, $1 \le k < n_i$
 $= 1$ $X_{n_i}^{(i)} \le x$.

We assume, without loss of generality, that $X_1^{(i)} < X_2^{(i)} < \ldots < X_{n_i}^{(i)}$.

To test the hypothesis that the population distribution functions are identical, Birnbaum and Hall introduced the statistics

$$D(n_1, n_2, ..., n_c) = \sup_{x,i,j} |F^{*(i)}(x) - F^{*(j)}(x)|,$$
 (1)

$$D^{+}(n_{1},n_{2}, ...,n_{c}) = \sup_{x,i < j} [F^{+(i)}(x) - F^{+(j)}(x)]$$
 (2)

i,j = 1,2, ..., c which are generalizations of the well known Kolmogorov-Smirnov two sample statistics D(m, n) and $D^{+}(m, n)$.

Under the null hypothesis

$$H_0: F^{(i)} = F^{(j)}$$
 i, j = 1,2, ..., c,

the exact small sample distributions of the statistics (1) and (2) were determined by Birnbaum-Hall using difference equations that could be solved recursively and tabled values of

$$P[D(n, n, n) \le r], P[D(n,n) \le r], P[D^+(n,n) \le r]$$

were developed for selected values of n between 1 and 40 and of r = k/n, $k = 1, 2, \ldots, n$.

¹Z. W. Birnbaum and R. A. Hall, "Small Sample Distributions for Multi-Sample Statistics of the Smirnov Type," Ann. Math. Stat., Vol. 31 (1960), pp. 710-720.

The limitations on the tables were for the most part imposed by the speed and capacity of the computer necessary to solve the difference equations. This paper utilizes some notions from graph theory, and in so doing, provides exact small sample distributions for more than three samples and for unequal sample sizes. Representative tables are included.

II. A GEOMETRICAL PERSPECTIVE

For a geometrical perspective of the problem we parallel in the next two paragraphs the development of Birnbaum-Hall.

The c-dimensional random variable $[F^{*(1)}(x), F^{*(2)}(x), \ldots, F^{*(c)}(x)]$, for fixed x, takes on values $(k_1/n_1, k_2/n_2, \ldots, k_c/n_c), k_i \in \{0, 1, \ldots, n_i\}$; $i = 1, 2, \ldots, c$, which lie in the unit hypercube. The transformation (x_1, x_2, \ldots, x_c) into (y_1, y_2, \ldots, y_c) defined by $y_i = n_i x_i$ maps the hypercube into a c-dimensional hypervolume with sides (n_1, n_2, \ldots, n_c) and the random c-vector into points (k_1, k_2, \ldots, k_c) with integer valued coordinates.

Under the null hypothesis H_0 : $F^{(i)} = F^{(j)}$ i, $j = 1, 2, \ldots$, c we may consider the c samples as coming from the same population with N! equally likely ways of drawing the combined sample of size $N = n_1 + n_2 + \ldots + n_c$. Now suppose the values of the combined ordered sample are located on the real line, and the null c-vector is used as a counter in the following manner: Starting at a value less than min $\{\chi_1^{(1)}, \chi_1^{(2)}, \ldots, \chi_1^{(c)}\}$ with the counter equal $(0, 0, \ldots, 0)$ and moving in the direction of increasing magnitude, the kth coordinate of the counter is incremented by one whenever a sample value $X_i^{(k)}$ is encountered. In so doing, a 1-1 correspondence between a path in c-space from $(0,0,\ldots,0)$ to (n_1,n_2,\ldots,n_c) and the c samples is defined.

The total number of such equally likely paths is the multinomial coefficient N!/ $(n_1!n_2! \dots n_c!)$. The critical region for the test will consist of a set of points R = $\{(k_1, k_2, \dots, k_c)\}$ such that any path from $(0,0,\dots,0)$ to (n_1,n_2,\dots,n_c) passing through R will cause H_0 to be rejected. If we denote the number of paths passing through R as $Q(n_1,n_2,\dots,n_c;R)$ then the probability of an error of the first kind

will be

$$\alpha = \frac{Q(n_1, n_2, ..., n_c; R)}{N!/(n_1!n_2! ... n_c!)}.$$

III. DETERMINATION OF EXACT SMALL SAMPLE DISTRIBUTION

Determination of the exact small sample distribution of (1) and (2) from this approach involves an enumerative process to determine $Q(n_1,n_2,\ldots,n_c;R)$. The set R defining the critical region suggests itself as $R = \{(k_1,k_2,\ldots,k_c) \mid \sup_{i,j} \mid n_j k_i - n_i k_j \mid n_j n_j r\}$ and

$$R^+ = \{(k_1, k_2, ..., k_c) \mid \sup_{i < j} (n_j k_i - n_i k_j) > n_i n_j r\}$$
 corresponding to

 $D(n_1, n_2, \ldots, n_c) > r$ and $D^+(n_1, n_2, \ldots, n_c) > r$ respectively. These regions of rejection are again extensions of the Kolmogorov-Smirnov procedure; however, there is no theoretical basis for their limitation to this form.

Considering the paths in c-space from $(0,0,\ldots,0)$ to (n_1,n_2,\ldots,n_c) as simple paths in a finite directed graph, (see e.g. Berge²) we have at our disposal a method of enumeration appearing in a paper by Pototchi³ which seems well suited to computer manipulation. This method, while conceptually analogous to the difference equation evaluated by Birnbaum-Hall, allows us to accomplish the enumeration necessary to evaluate $Q(n_1,n_2,\ldots,n_c;R)$ and to consider the number of samples c>3 as well as unequal sample sizes in an efficient and economical manner.

IV. TABLES

Table 1 describes the upper tail of the distribution of D, and contains the probabilities $P[D(n_1,n_2,n_3) \le r]$ for $5 \le n_1,n_2,n_3 \le 10$ starting at the largest increment of r = .1(.1)1.0 for which the tail probability equals or exceeds .20.

²C. Berge, The Theory of Graphs and its Applications, John Wiley & Sons, New York, 1962.

A. Pototchi, "A Simple Algorithm for Determining the Number of Paths in a Finite Graph," Economic Cyber Studies Res., Vol. 2 (1967), pp. 81-85.

Table 2 contains the probabilities $P[D(n,n,n,n) \le r]$ for n = 2(1)10 corresponding to values of r = k/n consistent with the differences $|F^{*(i)}(x) - F^{*(j)}(x)|$ which can occur.

The values in Table 1 for the triples (n,n,n), n=5(1)10 correspond to those computed by Birnbaum-Hall. The values in Table 2 for the 4-tuples (n,n,n,n), n=5,10 are consistent with those Monte Carloed by Gardner, 4 et al. Taken together, these provide additional verification of the accuracy of the tables.

⁴R. H. Gardner, J. E. Pinder III, and R. S. Wood, "Monte Carlo Estimation of Percentiles for the Multi-Smirnov Test," J. Statist. Comput. Simul., Vol. 10 (1980), pp. 243-249.

TABLE 1. QUANTILES OF THE BIRNBAUM-HALL, OR THREE SAMPLE SMIRNOV, STATISTIC FOR UNEQUAL SAMPLE SIZES.

ⁿ 1	n ₂	n ₃	r	$P[D(n_1,n_2,n_3) \leq r]$	n ₁	n ₂	n ₃	r	$P[D(n_1,n_2,n_3) \leq r]$
5	5	5	.6 .7 .8 .9	.81109 .81109 .97819 .97819 1.00000	5	6	6	.6 .7 .8 .9	.64253 .89712 .93216 .98982 1.00000
5	5	6	.6 .7 .8 .9	.67353 .85609 .94628 .98457	5	6	7	.6 .7 .8 .9	.70972 .87336 .95008 .99238 1.00000
5	5	7	.6 .7 .8 .9	.75909 .83058 .96542 .98776 1.00000	5	6	8	.6 .7 .8 .9	.70123 .88876 .95968 .99372 1.00000
5	5	8	.6 .7 .8 .9	.74472 .86485 .97582 .98947 1.00000	5	6	9	.6 .7 .8 .9	.72578 .91119 .96508 .99446 1.00000
5	5	9	.5 .6 .7 .8 .9	.46916 .80259 .88479 .98176 .99043	5	6	10	.6 .7 .8 .9	.75753 .92490 .96825 .99490 1.00000
5	5	10	.5 .6 .7 .8 .9	.52097 .83928 .89685 .98531 .99100 1.00000	5	7	7	.6 .7 .8 .9	.77668 .84966 .96547 .99462 1.00000
					5	7	8	.6 .7 .8 .9	.74794 .88248 .97350 .99578 1.00000

TABLE 1. (CONT'D)

n_1	n ₂	n ₃	r	$P[D(n_1,n_2,n_3) \leq r]$	<u>n</u> 1	n ₂	n ₃	r	$P[D(n_1,n_2,n_3) \leq r]$
5	7	9	.6 .7 .8 .9	.78143 .90138 .97793 .99642 1.00000	5	10	10	.5 .6 .7 .8 .9	.70953 .88851 .96359 .99235 .99870 1.00000
5	7	10	.5 .6 .7 .8 .9	.54453 .81483 .91269 .98047 .99680 1.00000	6	6	6	.6 .7 .8 .9	.68408 .93216 .93216 .99383 1.00000
5	8	8	.6 .7 .8 .9	.74256 .91251 .98065 .99685 1.00000	6	6	7	.6 .7 .8 .9	.74897 .91138 .94990 .99569 1.00000
5	8	9	.6 .7 .8 .9	.78949 .92949 .98457 .99745 1.00000	6	6	8	.6 .7 .8 .9	.74664 .90488 .95945 .99662 1.00000
5	8	10	.5 .6 .7 .8 .9	.62530 .81808 .93951 .98681 .99780 1.00000	6	6	9	.6 .7 .8 .9	.74555 .92900 .96485 .99712 1.00000
5	9	9	.5 .6 .7 .8 .9	.59437 .83481 .94532 .98821 .99803	6	6	10	.6 .7 .8 .9 1.0	.78119 .94384 .96806 .99739 1.00000
5	9	10	.5 .6 .7 .8 .9	.65118 .86208 .95464 .99031 .99836 1.00000	6	7	7	.5 .6 .7 .8 .9	.49330 .81149 .88975 .96517 .99719

TABLE 1. (CONT'D)

<u>n</u> 1	n ₂	n ₃		$P[D(n_1, n_2, n_3) \leq r]$	ⁿ 1	n ₂	n ₃	r	$P[D(n_1, n_2, n_3) \leq r]$
6	7	8	.6	.78637	6	9	10	.5	.67974
			.7	.90257				.6	.84898
			.8	.97319				.7	.95453
			.9	.99792				.8	.99011
			1.0	1.00000				.9	.99934
								1.0	1.00000
6	7	9	.6	.79154	6	10	10	.5	.70263
			.7	.92361				.6	.87756
			.8	.97765				. 7	.96382
			. 9	.99830				. 5	.99217
			1.0	1.00000				9	.99950
			1.0	1.0000				1.0	1.00000
6	7	10	.5	.57700	7	7	7	.5	.56209
•	•		.6	.82881	•	•	,	.6	.86823
			.7	.93638				.7	.86823
			.8	.98023				.8	.97750
			.9	.99851					
								.9	.99830
			1.0	1.00000				1.0	1.00000
6	8	8	.6	.78567	7	7	8	.5	.63985
			.7	.91100				. 6	.81671
			.8	.98035				.7	.89916
			.9	.99855				.8	.98371
			1.0	1.00000				.9	.99883
								1.0	1.00000
6	8	9	.5	.61580	7	7	9	.5	.61004
			.6	.80505				.6	.82936
			.7	.92848				.7	.91681
			.8	.98430				. 8	.98702
			.9	.99888				.9	.99909
			1.0	1.00000				1.0	1.00000
6	8	10	.5	.65020	7	7	10	.5	.62594
			.6	.83826				.6	.86663
			.7	.93879				.7	.92731
			.8	.98659				.8	.98888
			.9	.99906				.9	.99922
			1.0	1.00000				1.0	1.00000
6	9	9	.5	.65335	7	8	8	.6	.79392
			.6	.81993				. 7	.92685
			.7	.94488				.8	.98888
			.8	.98798				.9	.99924
			.9	.99918				1.0	1.00000
			1.0	1.00000				1.0	1.00000
			1.0	1.00000					

TABLE 1. (CONT'D)

n ₁	n ₂	n ₃	r	$P[D(n_1, n_2, n_3) \le r]$	n ₁	n ₂	n ₃	r	$P[D(n_1,n_2,n_3) \leq r]$
7	8	9	.5	.64705	8	8	10	.5	.74735
			.6	.82084		-		.6	.86362
			.7	.94230				.7	.96997
			.8	.99158				.8	.99601
			.9	.99944				.9	.99976
			1.0	1.00000				1.0	1.00000
			1.0	1.00000				1.0	1.00000
7	8	10	.5	.68971	8	9	9	.5	.65584
			.6	.85147	_	_	_	.6	.87789
			.7	.95131				.7	.97370
			.8	.99306				.8	.99659
			.9	.99955				.9	.99980
			1.0	1.00000				1.0	1.00000
			1.0	1.00000				1.0	1.00000
7	9	9	.5	.66124	8	9	10	.5	.71259
			.6	.84333				.6	.90160
			.7	.95640				.7	.97966
			.8	.99394				.8	.99745
			.9	.99962				.9	.99985
			1.0	1.00000				1.0	1.00000
7	9	10	.5	.69235	8	10	10	.5	.76924
			.6	.87042				.6	.92368
			.7	.96455				.7	.98495
			.8	.99523				.8	.99819
			.9	.99970				.9	.99990
			1.0	1.00000				1.0	1.00000
_			_						
7	10	10	.5	.72037	9	9	9	.5	.71542
			.6	.89658				.6	.91350
			.7	.97223				.7	.98248
			.8	.99641				.8	.99785
			.9	.99979				.9	.99988
			1.0	1.00000				1.0	1.00000
8	8	8	.6	.79392	9	9	10	.5	.76978
		-	.7	.95029	-		10	.6	
			.8	.99292				.7	.93369 .98714
			.9	.99954				.8	.99848
			1.0	1.00000				.9	
			1.0	1.00000				1.0	.99992 1.00000
								1.0	1.00000
8	8	9	.5	.67466	9	10	10	. 4	.53864
			.6	.83749				.5	.82180
			.7	.96289				.6	.95150
			.8	.99494				.7	.99104
			.9	.99968				.8	.99898
			1.0	1.00000				.9	.99995
								1.0	1.00000
								1.0	1.0000

TABLE 1. (CONT'D)

n ₁	n ₂	n ₃	r	$P[D(n_1,n_2,n_3) \leq r]$
10	10	10	. 4	.63715
			.5	.86930
			.6	.96645
			.7	.99411
			.8	.99936
			.9	.99997
			1.0	1 00000

TABLE 2. QUANTILES OF A FOUR SAMPLE SMIRNOV STATISTIC FOR EQUAL SAMPLE SIZES.

n ₁	n ₂	n ₃	n ₄	r	$P[D(n_1, n_2, n_3, n_4) \le r]$	n,	n a	2 n	3 ⁿ 4	r	$P[D(n_1, n_2, n_3, n_4) \le r]$
2	2	2	2	.5 1.0	.22857	8	8	8	8	.500	
				***	1.00000					.750	
										.875	
										1.000	
3	3	3	3	.333		9	9	9	9	.444	.56199
				. 666	6 .64286					.555	
				1.000	1.00000					.666	
										.777	
										.888	
										1.000	
4	4	4	4	.25	.00526	10	10	10	10	.5	.78000
				.50	.35707					.6	.93879
				.75	.87219					.7	.98874
				1.00	1.00000					.8	.99875
										.9	.99994
										1.0	1.00000
5	5	5	5	.2	.00068						
				. 4	.18184						
				.6	.69490						
				.8	.95982						
				1.0	1.00000						
6	6	6	6	.500							
				.666							
				.833	.98824						
				1.000	1.00000						
7	7	7	7	57 1	77075						
,	/	,	′	.571							
				.714							
				.857							
				1.000	1.00000						

DISTRIBUTION LIST

No. of Copies		No. of Copies	Organization
	Commander Defense Technical Info Center ATTN: DDC-DDA Cameron Station Alexandria, VA 22314	1	Commander US Army Electronics Research and Development Command Technical Support Activity ATTN: DELSD-L Fort Monmouth, N.J 07703
_	Commander US Army Materiel Development and Readiness Command ATTN: DRCDMD-ST 5001 Eisenhower Avenue Alexandria, VA 22333	-	Commander US Army Missile Command ATTN: DRSMI-R Redstone Arsenal, AL 35809 Commander
-	Commander US Army Armament Research and Development Command ATTN: DRDAR-TSS (2 cys) Dover, NJ 07801	1	US Army Missile Command ATTN: DRSMI-YDL Redstone Arsenal, AL 35809 Commander US Army Tank Automotive Rsch
_	Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-LEP-L, Tech Lib Rock Island, IL 61299	1	and Development Command ATTN: DRDTA-UL Warren, MI 48090 Director
	Director US Army ARRADCOM Benet Weapons Laboratory ATTN: DRDAR-LCB-TL Watervliet, NY 12189	,	US Army TRADOC Systems Analysis Activity ATTN: ATAA-SL, Tech Lib White Sands Missile Range NM 88002
	Commander US Army Aviation Research and Development Command ATTN: DRSAV-E P.O. Box 209 St. Louis, MO 61366	1	Commander US Army Research Office ATTN: DRXRO-MA, Dr. Robert Launer P.O. Box 12211 Research Triangle Park, NC 27709 Commander US Army Combat Development
	Director US Army Air Mobility Research and Development Laboratory Ames Research Center		Experimentation Command ATTN: ATEC-SA, Dr. Marion R. Bryson Fort Ord, CA 93941
1	Moffett Field, CA 94035 Commander US Army Communications Rsch and Development Command ATTN: DRDCO-PPA-SA Fort Monmouth, NJ 07703		Commander US Army Operational Test and Evaluation Agency ATTN: CSTE-STS, Mr. Fred C. McCoy 5600 Columbia Pike Falls Church, VA 22041

DISTRIBUTION LIST

No. of

Copies Organization

Aberdeen Proving Ground

Dir, USAMSAA

ATTN: DRXSY-D

DRXSY-MP, H. Cohen

Cdr, USATECOM

ATTN: DRSTE-TO-F

DRSTE-AD, Dr. Ted Trybul Dir, USACSL, Bldg. E3516, EA ATTN: DRDAR-CLB-PA

USER EVALUATION OF REPORT

Please take a few minutes to answer the questions below; tear out this sheet, fold as indicated, staple or tape closed, and place in the mail. Your comments will provide us with information for improving future reports.

1. BRL Report Number
2. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which report will be used.)
3. How, specifically, is the report being used? (Information source, design data or procedure, management procedure, source of ideas, etc.)
4. Has the information in this report led to any quantitative savings as far as man-hours/contract dollars saved, operating costs avoided, efficiencies achieved, etc.? If so, please elaborate.
5. General Comments (Indicate what you think should be changed to make this report and future reports of this type more responsive to your needs, more usable, improve readability, etc.)
6. If you would like to be contacted by the personnel who prepared this report to raise specific questions or discuss the topic, please fill in the following information.
Name:
Telephone Number:
Organization Address:

- FOLD HERE -

Director
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005



OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE. \$300

BUSINESS REPLY MAIL

FIRST CLASS PERMIT NO 12062 WASHINGTON, DC

POSTAGE WILL BE PAID BY DEPARTMENT OF THE ARMY

Director
US Army Ballistic Research Laboratory
ATTN: DRDAR-TSB
Aberdeen Proving Ground, MD 21005

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

